Determination of Sample Size for Higher Volatile Data using New Framework of Hybrid Box-Jenkins - GARCH: A Case Study on Gold Price

Siti Roslindar Yaziz[1], Roslinazairimah Zakaria[1] and Maizah Hura Ahmad[2]

[1] Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, Malaysia.
[2] Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Malaysia.

Abstract. The hybrid model of Box-Jenkins - GARCH has been shown to be a promising tool for forecasting higher volatile time series. In this study, the framework of determining the optimal sample size using the hybrid Box-Jenkins - GARCH is proposed for practical application in analysing and forecasting higher volatile data. The proposed framework is employed to daily world gold price series from year 1971 to 2013. The data is divided into 12 different sample sizes (from 30 to 10200). Each sample is tested using different combination of the hybrid Box-Jenkins - GARCH model. Our study shows that the optimal sample size to forecast gold price using the framework of the hybrid model is 1250 data of 5-year sample. Hence, the empirical results of model selection criteria and 1-step-ahead forecasting evaluations suggest that the latest 12.25% (5-year data) of 10200 data is sufficient enough to be employed in the hybrid Box-Jenkins - GARCH with similar forecasting performance as by using 41-year data.

Keyword: Box-Jenkins model, Box-Jenkins – GARCH, Gold price forecasting, Sample size

1. Introduction

The hybrid model of Box-Jenkins - GARCH is proven as the promising method to analyse and forecast a higher volatile data series such as gold price [1–6], electricity price [7,8], internet traffic [9] and traffic flow [10]. However, there is no discussion on the appropriate sample size using the hybrid model in the previous literature. Therefore, this paper is aimed to propose the framework using the Box-Jenkins - GARCH on how to determine the optimal sample size for forecasting purposes. According to Hyndman and Kostenko, the number of data required for any statistical model depends on at least two things: the number of model parameters to estimate and the amount of random variation in the data [11]. In other words, a reasonable approach to determine the appropriate sample size for forecasting model is to ensure that there is enough data to estimate the model and the model performs well out-of-sample evaluation.

In order to obtain a parsimonious estimated model, the Akaike Information Criteria (AIC) and the Schwarz Information Criterion (SIC) are applied. While, to evaluate the forecasting performance, the out-of-sample 1-step-ahead forecasting evaluations that are the mean square error (MSE), the root mean square error (RMSE) and the mean absolute error (MAE) are applied. As the sample size increasing, minimising the AIC is equivalent to minimising the out-of-sample 1-step-ahead MSE [12]. The method of the selection criteria and the forecasting evaluations are incorporated in the proposed framework in finding the optimal sample size. The proposed framework is illustrated using world daily gold price. To the best of our knowledge, this study is considered a pioneer in determining the optimal sample size for the hybrid model of Box-Jenkins - GARCH.